DOCUMENT RESUME

ED 478 199 TM 035 044

AUTHOR	Fenton, Ray
TITLE	How Have State Level Standards-Based Tests Related to Norm-Referenced Tests in Alaska?.
PUB DATE	2003-04-00
NOTE	12p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 21-25, 2003).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Academic Achievement; Educational Assessment; *Norm Referenced Tests; State Programs; *State Standards; *Test Results; *Testing Programs
IDENTIFIERS	Alaska; *Alaska Statewide Assessment Program; TerraNova Multiple Assessments

ABSTRACT

This overview of the Alaska system for test development, scoring, and reporting explored differences and similarities between norm-referenced and standards-based tests. The current Alaska testing program is based on legislation passed in 1997 and 1998, and is designed to meet the requirements of the federal No Child Left Behind Legislation. In 2002-2003, the Alaska benchmark Tests, given in grades 3, 6, and 8, are standards-based, while the Terra Nova Cat 6 tests, given in grades 4, 5, 7, and 9, have normative reporting. The Alaska High School Graduation Qualifying Examination also uses standards-based reporting. Participation is also required in the National Assessment of Educational Progress testing. The overall system was designed to be a hybrid of standardized and norm referenced tests. Available data do not allow a determination of the extent to which norm-referenced and performance-referenced tests in Alaska perform in the same way, but a quick look suggests that there is substantial similarity between the norm-referenced and standards-based tests. The items come from the same item pools and are highly similar for both tests. However, the differences in the percentages reaching cut scores on the various tests and the association of cut scores with performance expressed in terms of national norms raises some very real questions about the tests in terms of what should be expected of both students and tests. The discussion of what constitutes a valid measure of performance relative to standards and growth expectations has to be explored as part of an ongoing effort to find fairness. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 478 199

How Have State Level Standards-Based Tests Related to Norm-Referenced Tests in Alaska?

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1 ✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

R. Fenton

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Ray Fenton, Ph.D.
FentonResearch
8520 N. Coral Ridge Loop
Tucson, AZ 85704

FentonResearch@aol.com
(520) 229-8687

TM035044

A paper prepared for presentation as part of the symposium "Do Standards-Based Tests Differ Fundamentally from Norm Referenced Tests in Practice?" at the American Educational Research Association Convention held in Chicago in April 2003.

BEST COPY AVAILABLE

Author's Notes

This paper could not have been completed without the material assistance of the State of Alaska DEED that provided the information used to describe the current testing system and the data used in the 2001. There have been substantial changes in cut scores and content of the Alaska exams since 2000. It is strongly suggested that the interested reader look at the reports now available from the Alaska Department of Education and Early Development WWW site for the most current information on Alaska examinations and cut scores, www.deed.ak.us.org.

The tables in this paper were derived from the much more extensive analysis of the validity of the 2000 administration of the Alaska Benchmark and High School Graduation Examinations. The interested reader is referred to a paper presented at the AERA convention in Seattle in April 2001 for an extended discussion of the validity of the Alaska examination system (Stofflet 2001). This paper is available through the ERIC system.

Introduction

Alaska has a long history of attempting to use assessments to improve instruction. Students in Anchorage, Alaska were required to pass graduation exams for grade eight and high school as early as 1915. The State of Alaska developed a criterion referenced testing program in the early 1970s, a state-wide writing assessment in the early 1980s, and standards based assessment system in the late 1980s. The current State of Alaska student assessment system is based on laws enacted in 1997 and 1998 by the Alaska State Legislature to ensure accountability for Alaska Public Schools. The current program is designed to meet the requirements of the Federal No Child Left Behind Legislation.

The Alaska State Student Assessment System

The most up-to-date information on the Alaska State Student Assessment System is available at the State of Alaska Department of Education and Early Development World Wide Web site (WWW.eed.state.ak.us) and from publications such as Participation Guidelines for Alaska Students in Student Assessments (Alaska DEED, October 2001).

Configuration of the Alaska Statewide Student Assessment (2002-2003)

Grades	Assessments	Use
3 – 6 – 8	Alaska Benchmark Tests	Standards Based Reporting
4 – 5 – 7 - 9	Terra Nova Cat 6	Normative Reporting
10	Alaska High School Graduation Qualifying Examination	Standards Based Reporting

In addition to the tests above that are administered to all regular education students in public schools, participation is required in National Assessment of Educational Progress testing. Alternative and alternate assessment systems are provided for special education students. Alternate assessments can meet standards and lead to high school graduation. Alternative assessments are for severely disabled students who do not participate in programs that will result in meeting grade level performance standards or receiving a high school diploma.

The State of Alaska provides a variety of reports to parents, teachers and public school officials. In addition to these reports extensive information is provided to the public. The Alaska DEED WWW site provides district and school information on performance on both standards based and norm referenced tests as part of school and district report cards. The goal of the system is to fulfill the Federal Requirements of No Child Left Behind.

State Standards in Alaska

Information on Alaska content and performance standards is available for a variety of State of Alaska DEED publications and through the WWW site (WWW.eed.state.ak.us). Alaska educational standards are presented as content and performance standards. Standards include both a general statement of the content represented in the standard and an expanded description of what a student should be able to do to demonstrate that the standard has been met. For example, below are the main elements of the Alaska Mathematics content standards and the much more detailed estimation and computation performance standard for students who are aged eight to ten (Alaska DEED, 2000). Items to be included in grade 3 and up Alaska Benchmark tests are developed to demonstrate that a student is meeting grade level performance standards.

Mathematics Content Standards

A. A student should understand mathematical facts, concepts, principals and theories.

B. A student should understand and be able to select and use a variety of problem-solving strategies.

C. A student should understand and be able to form the appropriate methods to define and explain mathematical relationships.

D. A student should be able to use logic and reason to solve mathematical problems.

E. A student should be able to apply mathematical concepts and processes to situations within and outside of the school.

Estimation and Computation (Ages 8-10)

1) Describe and use a variety of estimation strategies including rounding to the appropriate place value, multiplying by the powers of 10, and using front-end estimation to check the reasonableness of solutions; (M.A. 3)

2) Recall and use basic multiplication and division facts orally, with paper and pencil without a calculator; (M.A. 3)

3) Add and subtract whole numbers and fractions with common denominators to 12 and decimals, including money amounts, using models and algorithms; (M.A. 3)

4) Multiply and divide multi-digit whole numbers by 2-digit numbers limiting the 2-digit divisors to those that end in 0; multiply and divide decimals that represent money by whole numbers; (M.A.3)

5) Find equivalent fractions; convert between fractions and mixed numbers; and (M.A.3)

6) Develop and interpret scales and scale models. (M.A.3)

Content Standards exist for English/Language Arts, Mathematics, Science, Geography, Government and Citizenship, History, Skills for a Healthy Life, Arts, World Languages, Technology, Employability, and Library/Information Literacy. Performance standards have been developed and accepted in Reading, Writing, and Mathematics. Science performance standards are being developed.

The Alaska Nexus between Norms Referenced and Standards Based Tests – Test Construction.

Alaska has relied heavily on the assistance of CTB/McGraw-Hill for technical advice and support in the development of the Alaska State Student Assessment. CTB/McGraw-Hill has been chosen repeatedly to develop assessments, provide testing services including materials distribution, collection and scoring, and to assist the state in reporting results.

The overall Alaska system was designed to be a hybrid of standardized and norm referenced tests to provide for the reporting of student performance at various levels in terms of both state standards and national norms. Initial agreements called for the inclusion of items in the state standards based Benchmark Assessments derived from the CTB Terra Nova item pool in order to allow the linking of norm referenced CTB CAT/Terra Nova and standards based Alaska Benchmark Exams in English/Language Arts and Math. Over the years, the links were extended from the grade 3, 6, and 8 Benchmark exams to include the grade 10 high school graduation exams. This requirement results in a substantial number of items on the Alaska State Tests being highly similar to the items in actual use in the nationally normed CTB Terra Nova/CAT tests.

The process of test construction for Alaska Benchmark and Graduation Qualifying Examination tests is simple and straightforward. Alaska identified the performance standards that were most important for testing. CTB/McGraw Hill provided thousands of items keyed to those standards from the Terra Nova item pool or custom written to reflect Alaska standards. Alaska educators and parents reviewed items and eliminated those thought either to be unfair for Alaska students because of some sort of face bias or not consistent with the standards.

CTB then selected items from the remaining pool, constructed tests of reasonably equivalent difficulty, and field tested the tests. After the initial tests were developed, additional "field test" items were included in tests to allow the ongoing development of additional test forms.

Multiple response formats are allowed. While most test items are traditional multiple choice items, some items allow students to show their work in mathematics and provide for short or extended written responses. Items are scored by CTB with trained evaluators making judgments about the number of points to be awarded to specific short and extended open response items.

The Alaska tests have the look and feel of modern norm referenced tests. The key to the valid standards based interpretation of performance is the connection between the test items and specific performance expectations for Alaska students. The entire test development process is designed to guarantee that the tests are measures of Alaska performance standards.

The Alaska Nexus Between Norm Referenced and Standards Based Tests – Test Scores and Judgments of Proficiency

The rhetoric of standards is never easy. On the one hand; parents, educators, and politicians all want all students to reach advanced performance levels represented by "high" or even "world class" standards. On the other hand, parents, educators, and politicians do not want to have their children and their schools judged as failures when they do not meet even "minimum performance standards." So those who set the actual standards have to walk a fine line.

Many states like Alaska have started with high standards that few students are able to achieve. With high stakes tests where individuals may be denied a high school diploma, the reality of setting reasonable performance standards quickly comes to the fore.

Alaska has set itself an additional burden by seeking to implement an articulated series of norm referenced and standards based tests that allow the tracking of student growth over time. The nature of such a system is that there must be compromises made that result in tests being able to measure ability over a fairly wide range of performances that are above and below grade level performance standards. The need for a range in difficulty in items increases as students increase in age and develop more sophisticated knowledge and skills.

The articulated system requires that tests be aligned to content standards to the extent that performances in the content areas tested are related from year-to-year and that the test items allow for "growth" in performance. This requirement takes the item selection process a step beyond the simple selection of items that are consistent with standards at grade level and makes standards based interpretation of results more complex.

The process of setting cut scores follows the CTB/McGraw Hill Book Marking procedure. This process has proven to meet the requirements for cut score/performance level setting in many states over the past twenty years. It makes use of panels of stake holders more or less familiar with the performance of local students at the grade level of the test. These experts examine actual test items ordered based on student performance and select a point in the order of items that reflects proficiency or "meeting standards."

The mechanics of the Book Marking process are simple. After a representative group of students are tested the performance on each item is tallied. Items on the test are then organized into a book in ascending order from the item that most students answered correctly to the item that most students failed to answer correctly.

Judges then go through the book, place a bookmark where they feel that minimum competency would be demonstrated, discuss their placement with other judges, come to a consensus as to where the mark should be placed, and then declare their choice.

Alaska has added one more review step in the most recent round of standard setting. Judges are provided with information on the percent of students that would be classified as a success based on the selected cut score. Judges then have a chance to reset the score with some direct knowledge of the impact of their standard on students and schools. It is felt that this additional round of standard setting increases the chance that the cut score would be fair and appropriate.

Alaskans feel that this standard setting process is a reasonable approach to building bridges between the limited ability of a group of test items to represent standards, the gulf that exists between the rhetoric of "high" and "world class standards" and the reality of student ability and experience.

As with most systems based on human judgments, there is an ample supply of critics who decry the political nature of the standard setting process or the absolute nature of decisions based on exams and cut scores.

To what extent do norm referenced and performance referenced tests in Alaska perform in the same way?

This is a good question that is currently impossible to answer with the available data. The State of Alaska DEED has not released information on the performance of individual students across the state to researchers in a way that will allow the examination of scores for individuals on norm referenced and performance based tests. Students are not tested by the state on NRT and CRT tests within the same year. Year-to-Year performance information linking growth on NRT and CRT scales is not generally available but could be derived from state data systems. As the State of Alaska establishes a history of student performance in a consolidated student database, it may be possible to do empirical studies of the validity of the assessment system. Within test administrations, it would now be possible to look at NRT and CRT performance interpretations through the examination of the information from the Terra Nova NRT items included in the test to assess the tests as measures of growth.

An early examination of student performance on the first version of the Alaska State High School Graduation Exam was undertaken in the Anchorage School District where both NRT and CRT information was available on 3,135 grade 10 students tested in 2000 (Stofflet, 2001). Based on the initial testing of sophomore students, 78% met Alaska standards in Reading, 51% met Alaska standards in Writing, and 36% met Alaska standards in Mathematics. Cut scores for the 2000 exam were set without any knowledge of student pass rates and the substantial divergence in pass rate in mathematics was consistent with the experience of many states that used the Book Mark procedure without knowledge of the consequences (Smiley, 2000).

Students in grades 3, 6, 8, and 10 were tested in April 2000 with the California Achievement Test (Fifth Edition) in late March. These scores were matched with the scores from Benchmark Examinations and the HSGQE given in early March. Normal Curve Equivalent Scores from CAT Total Reading, Total Language Arts, and Total Math were used to compute scale scores to correlate with standard scores derived from the Benchmark and High School Qualifying exams.[1]

The correlations and percent of variation were determined by conducting a series of linear regressions using the Statistical package for the Social Sciences (SPSS, 1999). Table 1 displays the correlation and percent of variance explained in Benchmark and HSGQE tests in 2000 by the NRT scores.

Table 1
Anchorage School District
Regression Analyses Predicting Alaska Test Scores from CAT 5 Scores
Spring 2000

Grade/Test	N	Correlation	Percent of Explained Variance
Grade 3			
Reading	3,806	.82	67%
Writing	3,808	.82	68%
Math	3,812	.78	61%
Grade 6			
Reading	3,863	.78	62%
Writing	3,863	.80	65%
Math	3,862	.78	69%
Grade 8			
Reading	3,539	.78	60%
Writing	3,542	.78	61%
Math	3,531	.84	71%
Grade 10			
Reading	2,724	.78	61%
Writing	2,171	.79	63%
Math	1,108	.83	69%

[1] See Stofflet 2002 for a detailed discussion of the procedures used.

The CAT normal curve equivalent scores that were associated with the Benchmark and HSGQ cut scores were examined to determine the CAT5 percentile scores that would be associated with the "passing levels" on the Alaska tests. Table 2 provides the Alaska Cut Score Scale Scores and associated percentile rank scores. The Book Marking procedures were repeated and some of the Alaska tests modified over the years since 2000. Reports from participants in that process indicate that the current tests and cut scores would produce a similar but much less drastic pattern of differences in the "difficulty level" of the cut scores.

Table 2
Anchorage School District
Correspondence between Passing Cut Points on the Benchmark Tests/HSGQE and California Achievement Test Scores
Spring 2000 Data

Grade/Test	Cut Score Scale Score	CAT5 NCE Score	CAT5 Percentile
Grade 3			
Reading	310	38.9	30
Writing	352	50.4	51
Math	322	44.3	39
Grade 6			
Reading	311	36.6	26
Writing	300	37.2	27
Math	329	49.3	49
Grade 8			
Reading	271	25.6	12
Writing	316	38.1	29
Math	376	61.9	72
Grade 10			
Reading	305	35.6	25
Writing	356	55.3	60
Math	383	68.4	81

Where does all this leave us?

This quick look at the Alaska system for test development, scoring, and reporting shows a substantial similarity between norm referenced and standards based tests. The items come from the same item pools and are highly similar for both types of tests. Items are selected to be consistent with what students are expected to learn by certain points in their careers and to provide enough diversity in item difficulty to assess year-to-year or point-to-point growth. Tests are constructed with item formats that will provide a

reasonably reliable indicator of performance. Group performance information is taken into account in the setting of final cut scores.

However, the differences in the percentages reaching cut scores on the various tests and the association of cut scores with performance expressed in terms of national norms raises some very real questions about the tests in terms of what should be expected of both students and tests. It is clear that the standards applied in 2000 were not based on consistent beliefs about acceptable performance. The discussion of what constitutes a valid measure of performance relative to standards and growth expectations has to be explored as part of an ongoing effort to find fairness. It is obvious that there is a need for ongoing studies of the validity of the overall assessment system.

Given the 2000 Alaska-Anchorage example, it is good to keep in mind the path suggested by Jaeger (1994) and to allow periodic review of performances relative to some indicator beyond of the internal indicators of test reliability and validity that are often cited as sufficient for justification of use of a given test.

> Whereas traditional validity standards might have been likened to truth in labeling laws, contemporary validity standards are more analogous to requirements for testing a new drug, with attention to the side effects as well as the intended benefits (p. 19).

This comment becomes even more salient when considered in terms of the current expectations set out in No Child Left Behind. The mandated\ that all students meet or exceed state standards by 2014 will put unrelenting pressure on schools and students. Annual or periodic review of the tests and the impact of the status and adequate yearly progress classifications need to be done to assure that the "side effects" are not causing harmful distortions in the educational system.

What will happen if it comes to pass that performance on the quasi-norm referenced tests developed to measure standards continues to be more or less normally distributed and the students who are low performers do not "rise up" to a level acceptable to those who desire performances that reflect "high" and "world class standards" for every student?

Brother, where art thou?

References

Alaska Department of Education and Early Development (2001). Participation Guidelines for Alaska Students in State Assessments 2001/2003. Alaska Department of Education and Early Development, Juneau. October 2001.

Alaska Department of Education and Early Development (2000).Alaska Standards: Content and Performance Standards for Alaska Students. Alaska Department of Education and Early Development, Juneau. February 2000.

Richard M. Jaeger (1994). Series Introduction. In G. Camilli and L.A. Shepard, MMSS: Methods for Identifying Biased Test Items, Volume 4 (pp. ix-xii). Thousand Oaks, CA: SAGE Publications, Inc.

Richard Smiley (2000). Private communication and report to the Alaska State Department of Education Technical Advisory Committee based on a telephone survey of state testing directors conducted by the Alaska Department of Education in 2002.

Fred Stofflet, Ray Fenton, and Tom Straugh (2001). Construct and Predictive Validity of the Alaska State High School Graduation Qualifying Examination: First Administration. A paper presented to the 2001 American Educational Research Association in Seattle WA in April 2001. This paper is available through ERIC.

Paper Presented at AERA Convention
April 2003

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

Educational Resources Information Center

REPRODUCTION RELEASE

(Specific Document)

TM035044

I. DOCUMENT IDENTIFICATION:

Title: How Have State Level Standards-Based Tests Related to Norm-Referenced Tests in Alaska?	
Author(s): Ray Fenton	
Corporate Source: Fenton Research 8520 N Coral Ridge Loop Tucson, AZ 85704	Publication Date: April 23, 2003

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

X

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please

Signature: Ray Fenton	Printed Name/Position/Title: Ray Fenton
Organization/Address: Fenton Research 8520 N Coral Ridge Loop Tucson, AZ 85704	Telephone: 707-830-3875 FAX: 520-229-8687 E-Mail Address: FentonResearch@aol.com Date: 5-5-03

(Over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse: **University of Maryland**
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Lab, Bldg 075
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Lab, Bldg 075
College Park, MD 20742
Attn: Acquisitions